Asymptotics for the length of the longest increasing subsequence of a binary Markov random word

Christian Houdré * Trevis J. Litherland †

August 27, 2012

Abstract

Let $(X_n)_{n\geq 0}$ be an irreducible, aperiodic, and homogeneous binary Markov chain and let LI_n be the length of the longest (weakly) increasing subsequence of $(X_k)_{1\leq k\leq n}$. Using combinatorial constructions and weak invariance principles, we present elementary arguments leading to a new proof that (after proper centering and scaling) the limiting law of LI_n is the maximal eigenvalue of a 2×2 Gaussian random matrix. In fact, the limiting shape of the RSK Young diagrams associated with the binary Markov random word is the spectrum of this random matrix.

AMS 2000 Subject Classification: 60C05, 60F05, 60F17, 60G15, 60G17, 05A16

Keywords: Longest increasing subsequence, Markov chains, Functional Central Limit Theorem, Random Matrices, Young diagrams

---

*Georgia Institute of Technology, School of Mathematics, Atlanta, Georgia, 30332-0160, USA, houdre@math.gatech.edu. Research supported in part by the NSA Grant H98230-09-1-0017.

†Georgia Institute of Technology, School of Mathematics, Atlanta, Georgia, 30332-0160, USA, trevisl@math.gatech.edu

1 Introduction

The identification of the limiting distribution of the length of the longest increasing subsequence of a random permutation or of a random word has attracted a lot of interest in the past decade, in particular in light of its connections with random matrices (see [1, 2, 3], [4], [6], [8], [12], [13, 14], [15], [17], [18]). For random words, both the iid uniform and non-uniform settings are understood, leading respectively to the maximal eigenvalue of a traceless (or generalized traceless) element of the Gaussian Unitary Ensemble (GUE) as limiting laws of LI_n. In a dependent framework, Kuperberg [16] conjectured that if the word is generated by an irreducible, doubly-stochastic, cyclic, Markov chain with state space an ordered m-letter alphabet, then the limiting distribution of the length LI_n is still that of the maximal eigenvalue of a traceless $m \times m$ element of the GUE. More generally, the conjecture asserts that the shape of the Robinson-Schensted-Knuth (RSK) Young diagrams associated with the Markovian random word is that of the joint distribution of the eigenvalues of a traceless $m \times m$ element of the GUE. For $m = 2$, Chistyakov and Götze [7] positively answered this conjecture, and in the present paper this result is rederived in an elementary way.

The precise class of homogeneous Markov chains with which Kuperberg's conjecture is concerned is more specific than the ones we shall study. The irreducibility of the chain is a basic property we certainly must demand: each letter has to occur at some point following the occurrence of any given letter. The cyclic (also called *circulant*) criterion, *i.e.*, the Markov transition matrix P has entries satisfying $p_{i,j} = p_{i+1,j+1}$, for $1 \leq i, j \leq m$ (where $m + 1 = 1$), ensures a uniform stationary distribution.

Let us also note that Kuperberg implicitly assumes the Markov chain to also be aperiodic. Indeed, the simple 2-state Markov chain for the letters α_1 and α_2 described by $\mathbb{P}(X_{n+1} = \alpha_i | X_n = \alpha_j) = 1$ for $i \neq j$, produces a sequence of alternating letters, so that LI_n is always either $n/2$ or $n/2+1$, for n even, and $(n+1)/2$, for n odd, and so has a degenerate limiting distribution. Even though this Markov chain is irreducible and cyclic, it is periodic.

By the end of this introduction, the reader might certainly have wondered how the binary results do get modified for ordered alphabets of arbitrary fixed size m. As shown in [10], for $m = 3$, Kuperberg's conjecture is indeed true. However, for $m \geq 4$, this is no longer the case; and some, but not all, cyclic Markov chains lead to a limiting law as in the iid uniform case.

2 Combinatorics

As in [9], one can express LI_n in a combinatorial manner. For convenience, this short section recapitulates that development.

Let $(X_n)_{n\geq 1}$ consist of a sequence of values taken from an m-letter ordered alphabet, $\mathcal{A}_m = \{\alpha_1 < \alpha_2 < \cdots < \alpha_m\}$. Let a_k^r be the number of occurrences of α_r among $(X_i)_{1\leq i\leq k}$. Each increasing subsequence of $(X_i)_{1\leq i\leq k}$ consists simply of consecutive identical values, with these values forming an increasing subsequence of α_r. Moreover, the number of occurrences of $\alpha_r \in \{\alpha_1, \ldots, \alpha_m\}$ among $(X_i)_{k+1\leq i\leq \ell}$, where $1 \leq k < \ell \leq n$, is simply $a_\ell^r - a_k^r$. The length of the longest increasing subsequence of $X_1, X_2, \ldots, X_n$ is thus given by

$$LI_n = \max_{\substack{0\leq k_1\leq\cdots \\ \leq k_{m-1}\leq n}} [(a_{k_1}^1 - a_0^1) + (a_{k_2}^2 - a_{k_1}^2) + \cdots + (a_n^m - a_{k_{m-1}}^m)], \tag{2.1}$$

i.e.,

$$LI_n = \max_{\substack{0\leq k_1\leq\cdots \\ \leq k_{m-1}\leq n}} [(a_{k_1}^1 - a_{k_1}^2) + (a_{k_2}^2 - a_{k_2}^3) + \cdots + (a_{k_{m-1}}^{m-1} - a_{k_{m-1}}^m) + a_n^m], \tag{2.2}$$

where $a_0^r = 0$.

For $i = 1, \ldots, n$ and $r = 1, \ldots, m-1$, let

$$Z_i^r = \begin{cases} 1, & \text{if } X_i = \alpha_r, \\ -1, & \text{if } X_i = \alpha_{r+1}, \\ 0, & \text{otherwise}, \end{cases} \tag{2.3}$$

and let $S_k^r = \sum_{i=1}^k Z_i^r$, $k = 1, \ldots, n$, with also $S_0^r = 0$. Then clearly $S_k^r = a_k^r - a_k^{r+1}$. Hence,

$$LI_n = \max_{\substack{0\leq k_1\leq\cdots \\ \leq k_{m-1}\leq n}} \{S_{k_1}^1 + S_{k_2}^2 + \cdots + S_{k_{m-1}}^{m-1} + a_n^m\}. \tag{2.4}$$

By the telescoping nature of the sum $\sum_{k=r}^{m-1} S_n^k = \sum_{k=r}^{m-1}(a_n^k - a_n^{k+1})$, we find that, for each $1 \leq r \leq m-1$, $a_n^r = a_n^m + \sum_{k=r}^{m-1} S_n^k$. Since $a_k^1, \ldots, a_k^m$ must evidently sum up to k, we have

$$
\begin{aligned}
n &= \sum_{r=1}^{m} a_n^r \\
&= \sum_{r=1}^{m-1} \left(a_n^m + \sum_{k=r}^{m-1} S_n^k \right) + a_n^m \\
&= \sum_{r=1}^{m-1} r S_n^r + m a_n^m.
\end{aligned}
$$

Solving for a_n^m gives us

$$
a_n^m = \frac{n}{m} - \frac{1}{m} \sum_{r=1}^{m-1} r S_n^r.
$$

Substituting into (2.4), we finally obtain

$$
LI_n = \frac{n}{m} - \frac{1}{m} \sum_{r=1}^{m-1} r S_n^r + \max_{\substack{0 \le k_1 \le \cdots \\ \le k_{m-1} \le n}} \{ S_{k_1}^1 + S_{k_2}^2 + \cdots + S_{k_{m-1}}^{m-1} \}. \tag{2.5}
$$

As emphasized in [9], (2.5) is of a *purely combinatorial nature or, in more probabilistic terms, is of a pathwise nature*. We now proceed to analyze (2.5) for a binary Markovian sequence.

3 Binary Markovian Alphabet

In the context of binary Markovian alphabets, $(X_n)_{n \ge 0}$ is described by the following transition probabilities between the two states (which we identify with the two letters α_1 and α_2): $\mathbb{P}(X_{n+1} = \alpha_2 | X_n = \alpha_1) = a$ and $\mathbb{P}(X_{n+1} = \alpha_1 | X_n = \alpha_2) = b$, where $0 < a + b < 2$. We later examine the degenerate cases $a = b = 0$ and $a = b = 1$. In keeping with the common usage within the Markov chain literature, we begin our sequence at $n = 0$, although our focus will be on $n \ge 1$. Denoting by (p_n^1, p_n^2) the vector describing the probability distribution on $\{\alpha_1, \alpha_2\}$ at time n, we have

$$\left(p_{n+1}^1, p_{n+1}^2\right) = \left(p_n^1, p_n^2\right) \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}. \tag{3.1}$$

The eigenvalues of the matrix in (3.1) are $\lambda_1 = 1$ and $-1 < \lambda_2 = 1 - a - b < 1$, with respective left eigenvectors $(\pi_1, \pi_2) = (b/(a+b), a/(a+b))$ and $(1, -1)$. Moreover, (π_1, π_2) is also the stationary distribution. Given any initial distribution (p_0^1, p_0^2), we find that

$$\left(p_n^1, p_n^2\right) = \left(\pi_1, \pi_2\right) + \lambda_2^n \frac{ap_0^1 - bp_0^2}{a+b} \left(1, -1\right) \to \left(\pi_1, \pi_2\right), \tag{3.2}$$

as $n \to \infty$, since $|\lambda_2| < 1$.

Our goal is now to use these probabilistic expressions to describe the random variables Z_k^1 and S_k^1 defined in the previous section. (We retain the redundant superscript "1" in Z_k^1 and S_k^1 in the interest of uniformity.)

Setting $\beta = ap_0^1 - bp_0^2$, we easily find that

$$\begin{aligned} \mathbb{E}Z_k^1 &= (+1)\left(\pi_1 + \frac{\beta}{a+b}\lambda_2^k\right) + (-1)\left(\pi_2 - \frac{\beta}{a+b}\lambda_2^k\right) \\ &= \frac{b-a}{a+b} + 2\frac{\beta}{a+b}\lambda_2^k, \end{aligned} \tag{3.3}$$

for each $1 \le k \le n$. Thus,

$$\mathbb{E}S_k^1 = \frac{b-a}{a+b}k + 2\left(\frac{\beta\lambda_2}{a+b}\right)\left(\frac{1-\lambda_2^k}{1-\lambda_2}\right), \tag{3.4}$$

and so $\mathbb{E}S_k^1/k \to (b-a)/(a+b)$, as $k \to \infty$.

Turning to the second moments of Z_k^1 and S_k^1, first note that $\mathbb{E}(Z_k^1)^2 = 1$, since $(Z_k^1)^2 = 1$ a.s. Next, we consider $\mathbb{E}Z_k^1 Z_\ell^1$, for $k < \ell$. Using the Markovian structure of $(X_n)_{n\ge 0}$, it quickly follows that

$$\begin{aligned} &\mathbb{P}((X_k, X_\ell) = (x_k, x_\ell)) \\ &= \begin{cases} \left(\pi_1 + \lambda_2^{\ell-k}\frac{a}{a+b}\right)\left(\pi_1 + \lambda_2^k \frac{\beta}{a+b}\right), & \text{if } (x_k, x_\ell) = (\alpha_1, \alpha_1), \\ \left(\pi_1 - \lambda_2^{\ell-k}\frac{b}{a+b}\right)\left(\pi_2 - \lambda_2^k \frac{\beta}{a+b}\right), & \text{if } (x_k, x_\ell) = (\alpha_1, \alpha_2), \\ \left(\pi_2 - \lambda_2^{\ell-k}\frac{a}{a+b}\right)\left(\pi_1 + \lambda_2^k \frac{\beta}{a+b}\right), & \text{if } (x_k, x_\ell) = (\alpha_2, \alpha_1), \\ \left(\pi_2 + \lambda_2^{\ell-k}\frac{b}{a+b}\right)\left(\pi_2 - \lambda_2^k \frac{\beta}{a+b}\right), & \text{if } (x_k, x_\ell) = (\alpha_2, \alpha_2). \end{cases} \end{aligned} \tag{3.5}$$

For simplicity, we will henceforth assume that our initial distribution is the stationary one, *i.e.*, $(p_0^1, p_0^2) = (\pi_1, \pi_2)$. (this assumption can be dropped as explained in the Concluding Remarks of [10]). Under this assumption, $\beta = 0$, $\mathbb{E}S_k^1 = k\mu$, where $\mu = \mathbb{E}Z_k^1 = (b-a)/(a+b)$, and (3.5) simplifies to

$$\begin{aligned}
&\mathbb{P}((X_k, X_\ell) = (x_k, x_\ell)) \\
&\quad = \begin{cases}
\left(\pi_1 + \lambda_2^{\ell-k}\frac{a}{a+b}\right)\pi_1, & \text{if } (x_k, x_\ell) = (\alpha_1, \alpha_1), \\
\left(\pi_1 - \lambda_2^{\ell-k}\frac{b}{a+b}\right)\pi_2, & \text{if } (x_k, x_\ell) = (\alpha_1, \alpha_2), \\
\left(\pi_2 - \lambda_2^{\ell-k}\frac{a}{a+b}\right)\pi_1, & \text{if } (x_k, x_\ell) = (\alpha_2, \alpha_1), \\
\left(\pi_2 + \lambda_2^{\ell-k}\frac{b}{a+b}\right)\pi_2, & \text{if } (x_k, x_\ell) = (\alpha_2, \alpha_2).
\end{cases}
\end{aligned} \tag{3.6}$$

We can now compute $\mathbb{E}Z_k^1 Z_\ell^1$:

$$\begin{aligned}
\mathbb{E}Z_k^1 Z_\ell^1 &= \mathbb{P}(Z_k^1 Z_\ell^1 = +1) - \mathbb{P}(Z_k^1 Z_\ell^1 = -1) \\
&= \mathbb{P}((X_k, X_\ell) \in \{(\alpha_1, \alpha_1), (\alpha_2, \alpha_2)\}) \\
&\qquad - \mathbb{P}((X_k, X_\ell) \in \{(\alpha_1, \alpha_2), (\alpha_2, \alpha_1)\}) \\
&= \left(\pi_1^2 + \lambda_2^{\ell-k}\frac{a}{a+b}\pi_1 + \pi_2^2 + \lambda_2^{\ell-k}\frac{b}{a+b}\pi_2\right) \\
&\qquad - \left(\pi_1\pi_2 - \lambda_2^{\ell-k}\frac{b}{a+b}\pi_2 + \pi_1\pi_2 - \lambda_2^{\ell-k}\frac{a}{a+b}\pi_1\right) \\
&= \left(\pi_1^2 + \pi_2^2 + \frac{2ab}{(a+b)^2}\lambda_2^{\ell-k}\right) - \left(2\pi_1\pi_2 - \frac{2ab}{(a+b)^2}\lambda_2^{\ell-k}\right) \\
&= \frac{(b-a)^2}{(a+b)^2} + \frac{4ab}{(a+b)^2}\lambda_2^{\ell-k}.
\end{aligned} \tag{3.7}$$

Hence, recalling that $\beta = 0$,

$$\begin{aligned}
\sigma^2 := \mathrm{Var} Z_k^1 &= 1 - \left(\frac{b-a}{a+b}\right)^2 \\
&= \frac{4ab}{(a+b)^2},
\end{aligned} \tag{3.8}$$

for all $k \geq 1$, and, for $k < \ell$, the covariance of Z_k^1 and Z_ℓ^1 is

$$\mathrm{Cov}(Z_k^1, Z_\ell^1) = \frac{(b-a)^2}{(a+b)^2} + \sigma^2 \lambda_2^{\ell-k} - \left(\frac{b-a}{a+b}\right)^2 = \sigma^2 \lambda_2^{\ell-k}. \tag{3.9}$$

Proceeding to the covariance structure of S_k^1, we first find that

$$\begin{aligned}
\mathrm{Var} S_k^1 &= \sum_{j=1}^{k} \mathrm{Var} Z_j^1 + 2 \sum_{j<\ell} \mathrm{Cov}(Z_j^1, Z_l^1) \\
&= \sigma^2 k + 2\sigma^2 \sum_{j<\ell} \lambda_2^{\ell-j} \\
&= \sigma^2 k + 2\sigma^2 \left(\frac{\lambda_2^{k+1} - k\lambda_2^2 + (k-1)\lambda_2}{(1-\lambda_2)^2} \right) \\
&= \sigma^2 \left(\frac{1+\lambda_2}{1-\lambda_2}\right) k + 2\sigma^2 \left(\frac{\lambda_2(\lambda_2^k - 1)}{(1-\lambda_2)^2}\right).
\end{aligned} \tag{3.10}$$

Next, for $k < \ell$, and using (3.9) and (3.10), the covariance of S_k^1 and S_ℓ^1 is given by

$$\begin{aligned}
\mathrm{Cov}(S_k^1, S_\ell^1) &= \sum_{i=1}^{k} \sum_{j=1}^{\ell} \mathrm{Cov}(Z_i^1, Z_j^1) \\
&= \sum_{i=1}^{k} \mathrm{Var} Z_i^1 + 2 \sum_{i<j<k} \mathrm{Cov}(Z_i^1, Z_j^1) + \sum_{i=1}^{k} \sum_{j=k+1}^{\ell} \mathrm{Cov}(Z_i^1, Z_j^1) \\
&= \mathrm{Var} S_k^1 + \sum_{i=1}^{k} \sum_{j=k+1}^{\ell} \mathrm{Cov}(Z_i^1, Z_j^1) \\
&= \mathrm{Var} S_k^1 + \sigma^2 \left(\frac{\lambda_2(1-\lambda_2^k)(1-\lambda_2^{\ell-k})}{(1-\lambda_2)^2} \right) \\
&= \sigma^2 \left(\left(\frac{1+\lambda_2}{1-\lambda_2}\right) k - \frac{\lambda_2(1-\lambda_2^k)(1+\lambda_2^{\ell-k})}{(1-\lambda_2)^2} \right).
\end{aligned} \tag{3.11}$$

From (3.10) and (3.11) we see that, as $k \to \infty$,

$$\frac{\mathrm{Var} S_k^1}{k} \to \sigma^2 \left(\frac{1+\lambda_2}{1-\lambda_2}\right), \tag{3.12}$$

and, moreover, as $k \wedge \ell \to \infty$,

$$\frac{\mathrm{Cov}(S_k^1, S_\ell^1)}{(k \wedge \ell)} \to \sigma^2 \left(\frac{1+\lambda_2}{1-\lambda_2} \right). \tag{3.13}$$

When $a = b$, $\mathbb{E} S_k^1 = 0$, and in (3.12) the asymptotic variance becomes

$$\begin{aligned} \frac{\mathrm{Var} S_k^1}{k} &\to \frac{4a^2}{(2a)^2} \left(\frac{1+(1-2a)}{1-(1-2a)} \right) \\ &= \frac{1}{a} - 1. \end{aligned}$$

For a small, we have a "lazy" Markov chain, that is, a Markov chain which tends to remain in a given state for long periods of time. In this regime, the random variable S_k^1 has long periods of increase followed by long periods of decrease. In this way, linear asymptotics of the variance with large constants occur. If, on the other hand, a is close to 1, the Markov chain rapidly shifts back and forth between α_1 and α_2, and so the constant associated with the linearly increasing variance of S_k^1 is small.

As in [9], Brownian functionals play a central rôle in describing the limiting distribution of LI_n. To move towards a Brownian functional expression for the limiting law of LI_n, define the polygonal function

$$\hat{B}_n(t) = \frac{S_{[nt]}^1 - [nt]\mu}{\sigma\sqrt{n(1+\lambda_2)/(1-\lambda_2)}} + \frac{(nt-[nt])(Z_{[nt]+1}^1 - \mu)}{\sigma\sqrt{n(1+\lambda_2)/(1-\lambda_2)}}, \tag{3.14}$$

for $0 \le t \le 1$. In our finite-state, irreducible, aperiodic, stationary Markov chain setting, we may conclude that $\hat{B}_n \Rightarrow B$, as desired. (See, for example, the more general settings for Gordin's martingale approach to dependent invariance principles, and the stationary ergodic invariance principle found in Theorem 19.1 of [5].)

Turning now to LI_n, we see that for the present 2-letter situation, (2.5) simply becomes

$$LI_n = \frac{n}{2} - \frac{1}{2} S_n^1 + \max_{1 \le k \le n} S_k^1.$$

To find the limiting distribution of LI_n from this expression, recall that $\pi_1 = b/(a+b)$, $\pi_2 = a/(a+b)$, $\mu = \pi_1 - \pi_2 = (b-a)/(a+b)$, $\sigma^2 =$

$4ab/(a+b)^2$, and that $\lambda_2 = 1 - a - b$. Define $\pi_{max} = \max\{\pi_1, \pi_2\}$ and $\tilde{\sigma}^2 = \sigma^2(1+\lambda_2)/(1-\lambda_2)$. Rewriting (3.14) as

$$\hat{B}_n(t) = \frac{S^1_{[nt]} - [nt]\mu}{\tilde{\sigma}\sqrt{n}} + \frac{(nt - [nt])(Z^1_{[nt]+1} - \mu)}{\tilde{\sigma}\sqrt{n}},$$

LI_n becomes

$$\begin{aligned} LI_n &= \frac{n}{2} - \frac{1}{2}\left(\tilde{\sigma}\sqrt{n}\hat{B}_n(1) + \mu n\right) + \max_{0\le t\le 1}\left(\tilde{\sigma}\sqrt{n}\hat{B}_n(t) + \mu nt\right) \\ &= n\pi_2 - \frac{1}{2}\left(\tilde{\sigma}\sqrt{n}\hat{B}_n(1)\right) + \max_{0\le t\le 1}\left(\tilde{\sigma}\sqrt{n}\hat{B}_n(t) + (\pi_1 - \pi_2)nt\right) \\ &= n\pi_{max} - \frac{1}{2}\left(\tilde{\sigma}\sqrt{n}\hat{B}_n(1)\right) \\ &\quad + \max_{0\le t\le 1}\left(\tilde{\sigma}\sqrt{n}\hat{B}_n(t) + (\pi_1 - \pi_2)nt - (\pi_{max} - \pi_2)n\right). \end{aligned} \tag{3.15}$$

This immediately gives

$$\begin{aligned} \frac{LI_n - n\pi_{max}}{\tilde{\sigma}\sqrt{n}} &= -\frac{1}{2}\hat{B}_n(1) \\ &\quad + \max_{0\le t\le 1}\left(\hat{B}_n(t) + \frac{\sqrt{n}}{\tilde{\sigma}}((\pi_1 - \pi_2)t - (\pi_{max} - \pi_2))\right). \end{aligned} \tag{3.16}$$

Let us examine (3.16) on a case-by-case basis. First, if $\pi_{max} = \pi_1 = \pi_2 = 1/2$, *i.e.*, if $a = b$, then $\sigma = 1$ and $\tilde{\sigma} = (1-a)/a$, and so (3.16) becomes

$$\frac{LI_n - n/2}{\sqrt{(1-a)n/a}} = -\frac{1}{2}\hat{B}_n(1) + \max_{0\le t\le 1}\hat{B}_n(t). \tag{3.17}$$

Then, by the Invariance Principle and the Continuous Mapping Theorem,

$$\frac{LI_n - n/2}{\sqrt{(1-a)n/a}} \Rightarrow -\frac{1}{2}B(1) + \max_{0\le t\le 1}B(t). \tag{3.18}$$

Next, if $\pi_{max} = \pi_2 > \pi_1$, (3.16) becomes

$$\frac{LI_n - n\pi_{max}}{\tilde{\sigma}\sqrt{n}} = -\frac{1}{2}\hat{B}_n(1) + \max_{0\le t\le 1}\left(\hat{B}_n(t) - \frac{\sqrt{n}}{\tilde{\sigma}}(\pi_{max} - \pi_1)t\right). \tag{3.19}$$

On the other hand, if $\pi_{max} = \pi_1 > \pi_2$, (3.16) becomes

$$\begin{aligned}\frac{LI_n - n\pi_{max}}{\tilde{\sigma}\sqrt{n}} &= -\frac{1}{2}\hat{B}_n(1) \\ &\quad + \max_{0\le t\le 1}\left(\hat{B}_n(t) - \frac{\sqrt{n}}{\tilde{\sigma}}(\pi_{max} - \pi_2)(1-t)\right) \\ &= \frac{1}{2}\hat{B}_n(1) \\ &\quad + \max_{0\le t\le 1}\left(\hat{B}_n(t) - \hat{B}_n(1) - \frac{\sqrt{n}}{\tilde{\sigma}}(\pi_{max} - \pi_2)(1-t)\right).\end{aligned} \tag{3.20}$$

In both (3.19) and (3.20) we have a term in our maximal functional which is linear in t or $1-t$, with a negative slope. We now show, in an elementary fashion, that in both cases, as $n \to \infty$, the maximal functional goes to zero in probability.

Consider first (3.19). Let $c_n = \sqrt{n}(\pi_{max} - \pi_1)/\tilde{\sigma} > 0$, and for any $c > 0$, let $M_c = \max_{0\le t\le 1}(B(t) - ct)$, where $B(t)$ is a standard Brownian motion. Now for n large enough,

$$\hat{B}_n(t) - ct \ge \hat{B}_n(t) - c_n t$$

a.s., for all $0 \le t \le 1$. Then for any $z > 0$, and n large enough,

$$\mathbb{P}(\max_{0\le t\le 1}(\hat{B}_n(t) - c_n t) > z) \le \mathbb{P}(\max_{0\le t\le 1}(\hat{B}_n(t) - ct) > z), \tag{3.21}$$

and so by the Invariance Principle and the Continuous Mapping Theorem,

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{P}(\max_{0\le t\le 1}(\hat{B}_n(t) - c_n t) > z) &\le \lim_{n\to\infty} \mathbb{P}(\max_{0\le t\le 1}(\hat{B}_n(t) - ct) > z) \\ &= \mathbb{P}(M_c > z).\end{aligned} \tag{3.22}$$

Now, as is well-known, $\mathbb{P}(M_c > z) \to 0$ as $c \to \infty$. One can confirm this intuitive fact with the following simple argument. For $z > 0$, $c > 0$, and $0 < \varepsilon < 1$, we have that

$$\begin{aligned}
\mathbb{P}(M_c > z) &\leq \mathbb{P}(\max_{0\leq t\leq\varepsilon}(B(t) - ct) > z) + \mathbb{P}(\max_{\varepsilon<t\leq 1}(B(t) - ct) > z) \\
&\leq \mathbb{P}(\max_{0\leq t\leq\varepsilon} B(t) > z) + \mathbb{P}(\max_{\varepsilon<t\leq 1}(B(t) - c\varepsilon) > z) \\
&\leq \mathbb{P}(\max_{0\leq t\leq\varepsilon} B(t) > z) + \mathbb{P}(\max_{0<t\leq 1} B(t) > c\varepsilon + z) \\
&= 2\left(1 - \Phi\left(\frac{z}{\sqrt{\varepsilon}}\right)\right) + 2\left(1 - \Phi(c\varepsilon + z)\right). \qquad (3.23)
\end{aligned}$$

But, as c and ε are arbitrary, we can first take the limsup of (3.23) as $c \to \infty$, and then let $\varepsilon \to 0$, proving the claim.

We have thus shown that

$$\limsup_{n\to\infty} \mathbb{P}(\max_{0\leq t\leq 1}(\hat{B}_n(t) - c_n t) > z) \leq 0,$$

and since the functional clearly is equal to zero when $t = 0$, we have

$$\max_{0\leq t\leq 1}(\hat{B}_n(t) - c_n t) \xrightarrow{\mathbb{P}} 0, \qquad (3.24)$$

as $n \to \infty$. Thus, by the Continuous Mapping Theorem, and the Converging Together Lemma, we obtain the weak convergence result

$$\frac{LI_n - n\pi_{max}}{\tilde{\sigma}\sqrt{n}} \Rightarrow -\frac{1}{2}B(1). \qquad (3.25)$$

Lastly, consider (3.20). Here we need simply note the following equality in law, which follows from the stationary and Markovian nature of the underlying sequence $(X_n)_{n\geq 0}$:

$$\begin{aligned}
&\hat{B}_n(t) - \hat{B}_n(1) - \frac{\sqrt{n}}{\tilde{\sigma}}(\pi_{max} - \pi_2)(1 - t) \\
&\qquad \stackrel{\mathcal{L}}{=} -\hat{B}_n(1 - t) - \frac{\sqrt{n}}{\tilde{\sigma}}(\pi_{max} - \pi_2)(1 - t), \qquad (3.26)
\end{aligned}$$

for $t = 0, 1/n, \ldots, (n-1)/n, 1$. With a change of variables ($u = 1 - t$), and noting that $B(t)$ and $-B(t)$ are equal in law, our previous convergence result (3.24) implies that

$$\max_{0 \le t \le 1} (\hat{B}_n(t) - \hat{B}_n(1) - c_n(1-t)) \stackrel{\mathcal{L}}{=} \max_{0 \le u \le 1} (-\hat{B}_n(u) - c_n u) \stackrel{\mathbb{P}}{\to} 0, \qquad (3.27)$$

as $n \to \infty$. Our limiting functional is thus of the form

$$\frac{LI_n - n\pi_{max}}{\tilde{\sigma}\sqrt{n}} \Rightarrow \frac{1}{2} B(1). \qquad (3.28)$$

Since $B(1)$ is simply a standard normal random variable, the different signs in (3.25) and (3.28) are inconsequential.

Finally, consider the degenerate cases. If either $a = 0$ or $b = 0$, then the sequence $(X_n)_{n \ge 0}$ will be a.s. constant, regardless of the starting state, and so $LI_n \sim n$. On the other hand, if $a = b = 1$, then the sequence oscillates back and forth between α_1 and α_2, so that $LI_n \sim n/2$. Combining these trivial cases with the previous development, gives:

Theorem 3.1 *Let $(X_n)_{n \ge 0}$ be a 2-state Markov chain, with $\mathbb{P}(X_{n+1} = \alpha_2 | X_n = \alpha_1) = a$ and $\mathbb{P}(X_{n+1} = \alpha_1 | X_n = \alpha_2) = b$. Let the law of X_0 be the invariant distribution $(\pi_1, \pi_2) = (b/(a+b), a/(a+b))$, for $0 < a + b \le 2$, and be $(\pi_1, \pi_2) = (1, 0)$, for $a = b = 0$. Then, for $a = b > 0$,*

$$\frac{LI_n - n/2}{\sqrt{n}} \Rightarrow \sqrt{\frac{1-a}{a}} \left(-\frac{1}{2} B(1) + \max_{0 \le t \le 1} B(t) \right), \qquad (3.29)$$

where $(B(t))_{0 \le t \le 1}$ is a standard Brownian motion. For $a \ne b$ or $a = b = 0$,

$$\frac{LI_n - n\pi_{max}}{\sqrt{n}} \Rightarrow N(0, \tilde{\sigma}^2/4), \qquad (3.30)$$

with $\pi_{max} = \max\{\pi_1, \pi_2\}$, and where $N(0, \tilde{\sigma}^2/4)$ is a centered normal random variable with variance $\tilde{\sigma}^2/4 = ab(2-a-b)/(a+b)^3$, for $a \ne b$, and $\tilde{\sigma}^2 = 0$, for $a = b = 0$. (If $a = b = 1$, or $\tilde{\sigma}^2 = 0$, then the distributions in (3.29) *and* (3.30)*, respectively, are understood to be degenerate at the origin.)*

To extend this result to the entire RSK Young diagrams, let us introduce the following notation. By

$$(Y_n^{(1)}, Y_n^{(2)}, \ldots, Y_n^{(k)}) \Rightarrow (Y_\infty^{(1)}, Y_\infty^{(2)}, \ldots, Y_\infty^{(k)}) \tag{3.31}$$

we shall indicate the weak convergence of the *joint* law of the k-vector $(Y_n^{(1)}, Y_n^{(2)}, \ldots, Y_n^{(k)})$ to that of $(Y_\infty^{(1)}, Y_\infty^{(2)}, \ldots, Y_\infty^{(k)})$, as $n \to \infty$. Since LI_n is the length of the top row of the associated Young diagrams, the length of the second row is simply $n - LI_n$. Denoting the length of the i^{th} row by R_n^i, (3.31), together with an application of the Cramér-Wold Theorem, recovers the result of Chistyakov and Götze [7] as part of the following easy corollary, which is in fact equivalent to Theorem 3.1:

Corollary 3.1 *For the sequence in Theorem 3.1, if $a = b > 0$, then*

$$\left(\frac{R_n^1 - n/2}{\sqrt{n}}, \frac{R_n^2 - n/2}{\sqrt{n}} \right) \Rightarrow Y_\infty := (R_\infty^1, R_\infty^2), \tag{3.32}$$

where the law of Y_∞ is supported on the 2^{nd} main diagonal of $\mathbb{R}^2$, and with

$$R_\infty^1 \stackrel{\mathcal{L}}{=} \sqrt{\frac{1-a}{a}} \left(-\frac{1}{2} B(1) + \max_{0 \le t \le 1} B(t) \right).$$

If $a \neq b$ or $a = b = 0$, then setting $\pi_{min} = \min\{\pi_1, \pi_2\}$, we have

$$\left(\frac{R_n^1 - n\pi_{max}}{\sqrt{n}}, \frac{R_n^2 - n\pi_{min}}{\sqrt{n}} \right) \Rightarrow N((0,0), \tilde{\Sigma}), \tag{3.33}$$

where $\tilde{\Sigma}$ is the covariance matrix

$$(\tilde{\sigma}^2/4) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},$$

where $\tilde{\sigma}^2 = 4ab(2 - a - b)/(a + b)^3$, for $a \neq b$, and $\tilde{\sigma}^2 = 0$, for $a = b = 0$.

Remark 3.1 *The joint distributions in* (3.32) *and* (3.33) *are of course degenerate, in that the sum of the two components is a.s. identically zero in each case. In* (3.32)*, the density of the first component of R_∞ is easy to find, and is given by (e.g., see [11])*

$$f(y) = \frac{16}{\sqrt{2\pi}} \left(\frac{a}{1-a} \right)^{3/2} y^2 e^{-2ay^2/(1-a)}, \qquad y \ge 0. \tag{3.34}$$

As in Chistyakov and Götze [7], (3.32) can then be stated as: For any bounded, continuous function $g : \mathbb{R}^2 \to \mathbb{R}$,

$$\lim_{n\to\infty} \left(g \left(\frac{R_n^1 - n/2}{\sqrt{(1-a)n/a}}, \frac{R_n^2 - n/2}{\sqrt{(1-a)n/a}} \right) \right)$$
$$= 2\sqrt{2\pi} \int_0^\infty g(x, -x) \phi_{GUE,2}(x, -x) dx,$$

where $\phi_{GUE,2}$ is the density of the eigenvalues of the 2×2 GUE, and is given by

$$\phi_{GUE,2}(x_1, x_2) = \frac{1}{\pi}(x_1 - x_2)^2 e^{-(x_1^2 + x_2^2)}.$$

To see the GUE connection more explicitly, consider the 2×2 traceless GUE matrix

$$M_0 = \begin{pmatrix} X_1 & Y + iZ \\ Y - iZ & X_2 \end{pmatrix},$$

where X_1, X_2, Y, and Z are centered, normal random variables. Since Corr $(X_1, X_2) = -1$, the largest eigenvalue of M_0 is

$$\lambda_{1,0} = \sqrt{X_1^2 + Y^2 + Z^2},$$

almost surely, so that $\lambda_{1,0}^2 \sim \chi_3^2$ if Var $X_1 =$ Var $Y =$ Var $Z = 1$. Hence, up to a scaling factor, the density of $\lambda_{1,0}$ is given by (3.34). Next, let us perturb M_0 to

$$M = \alpha G I + \beta M_0,$$

where α and β are constants, G is a standard normal random variable independent of M_0, and I is the identity matrix. The covariance of the diagonal elements of M is then computed to be $\rho := \alpha^2 - \beta^2$. Hence, to obtain a desired value of ρ, we may take $\alpha = \sqrt{(1+\rho)/2}$ and $\beta = \sqrt{(1-\rho)/2}$. Clearly, the largest eigenvalue of M can then be expressed as

$$\lambda_1 = \sqrt{\frac{1+\rho}{2}} G + \sqrt{\frac{1-\rho}{2}} \lambda_{1,0}. \tag{3.35}$$

At one extreme, $\rho = -1$, we recover $\lambda_1 = \lambda_{1,0}$. At the other extreme, $\rho = 1$, we obtain $\lambda_1 = Z$. Midway between these two extremes, at $\rho = 0$, we have a standard GUE matrix, so that

$$\lambda_1 = \sqrt{\frac{1}{2}} \left(G + \lambda_{1,0} \right).$$

References

[1] Baik, J., Deift, P., and Johansson, K., "On the distribution of the length of the longest increasing subsequence of random permutations," *J. Amer. Math. Soc.*, vol. 12, no. 4, pp. 1119-1178, 1999.

[2] Baik, J. Deift, P., and Johansson, K., "On the distribution of the length of the second row of a Young diagram under Plancherel measure," *Geom. Funct. Anal.*, vol. 10, no. 4, pp. 702-731, 2000.

[3] Baik, J. Deift, P., and Johansson, K., "Addendum to: "On the distribution of the length of the second row of a Young diagram under Plancherel measure"," *Geom. Funct. Anal.*, vol. 10, no. 6, pp. 1606-1607, 2000.

[4] Baryshnikov, Y., "GUEs and queues," *Probab. Theory Related Fields*, vol. 119, no. 2, pp. 256-274, 2001.

[5] Billingsley, P., *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics, New York: John Wiley and Sons Inc., second ed., 1999. A Wiley-Interscience Publication.

[6] Borodin, A., Okounkov, A., and Olshanski, G, "Asymptotics of Plancherel measures for symmetric groups," *J. Amer. Math. Soc.*, vol. 13, no. 3, pp. 481-515 (electronic), 2000.

[7] Chistyakov, G.P. and Götze, F., "Distribution of the shape of Markovian random words," *Probab. Theory Related Fields*, vol. 129, no. 1, pp. 18-36, 2004.

[8] Gravner, J., Tracy, C., and Widom, H., "Limit theorems for height fluctuations in a class of discrete space and time growth models," *J. Statist. Phys.*, vol. 102, no. 5-6, pp. 1085-1132, 2001.

[9] Houdré C. and Litherland, T., "On the longest increasing subsequence for finite and countable alphabets," in High Dimensional Probability V: The Luminy Volume (Beachwood, Ohio, USA: Institute of Mathematical Statistics), 185-212, 2009.

[10] Houdré C. and Litherland, T., On the limiting shape of Young diagrams associated with Markov random words. ArXiv #math.Pr/1110.4570, 2011.

[11] Houdré C., Lember, J., and Matzinger, H., "On the longest common increasing binary subsequence," *C. R. Acad. Sci. Paris*, vol. 343, no. 9, pp. 589-594, 2006.

[12] Houdré C. and Xu, H., "On the limiting shape of Young diagrams associated with inhomogeneous random words. To appear: HDP VI, 2013.

[13] Its, A.R., Tracy, C. A., and Widom, H., "Random words, Toeplitz determinants, and integrable systems. I," in *Random matrix models and their applications*, vol. 40 of *Math. Sci. Res. Inst. Publ.*, pp. 245-258, Cambridge: Cambridge Univ. Press, 2001.

[14] Its, A.R., Tracy, C. A., and Widom, H., "Random words, Toeplitz determinants, and integrable systems. II," *Phys. D*, vol 152/153, pp. 199-224, 2001. Advances in Nonlinear Mathematics and Science.

[15] Johansson, K., "Discrete orthogonal polynomial ensembles and the Plancherel measure," *Ann. of Math. (2)*, vol. 153, no.1, pp. 259-296, 2001.

[16] Kuperberg, G., "Random words, quantum statistics, central limits, random matrices," *Methods. Appl. Anal.*, vol. 9, no. 1, pp. 99-118, 2002.

[17] Okounkov, A., "Random matrices and random permutations," *Internat. Math. Res. Notices.*, , no. 20, pp. 1043-1095, 2000.

[18] Tracy, C. A. and Widom, H., "On the distributions of the lengths of the longest monotone subsequences in random words," *Probab. Theory Related Fields*, vol. 119, no. 3, pp. 350-380, 2001.